Muthanna Journal of Engineering and Technology, Vol. (12), Issue (2), Year (2024)



Muthanna Journal of Engineering and Technology

Website: https://muthjet.mu.edu.iq/

Submitted 27 June, Accepted 8 July 2024, Published online 12 July 2024



A comparison study on CNN-based brain tumor detection systems: proposed vs. pretrained models

Ahmed Saaudi^a, Riyadh Mansoor^a and Salah Alheejawi^b

^aElectronics and Communication Engineering, Al-Muthanna University, Samawah, Iraq

^bDepartment of Computer Techniques Engineering, College of Information Technology, Imam Ja'afar Al-Sadiq University, Al-Muthanna, Iraa

*Corresponding author E-mail: <u>ahmed.saaudi@mu.edu.iq</u>

DOI:10.52113/3/eng/mjet/2024-12-02/01-08

Abstract

A brain tumor is a serious disease that requires a talented specialist to differentiate the tumor types (benign or malignant) accurately in the early stages. Artificial intelligent (AI) can participate in facilitating the specialists' task by performing deep learning algorithms on MRI-based images to achieve an accurate decision. There are many pretrained models that are developed based on deep learning algorithms to tackle brain tumor identification issue. In this work, the weights of three pretrained models: Xception, Inception-resnet50, and VGG 16 are adapted to develop brain-tumor detection systems. The structure of each system is upgraded by adding an input layer and two dense layers of 32, and 16 nodes, respectively, with one output layer to classify the input samples, (MRI of brain tumor). Later, a comprehensive comparison is conducted to evaluate the behavior of each model according to the ability to identify the tumor type (Healthy or malignant). The comparison study reveals the superiority of the VGG16 model in terms of accuracy. Moreover, the structure of the VGG16 model presents less complexity regarding the number of CNN layers and training parameters. Reducing the complexity participates in saving the consuming energy and reducing the execution time. The latter is crucial since it helps specialists to identify the tumor type easily and in, relatively, less time. The main aim of this study is to develop a highly accurate and less complex CNN-based model to recognize the brain tumor type. A model of three CNN-based layers with 719281 trainable parameters is suggested. The proposed model shows 96% accuracy in approximately 8500 seconds. Even though the accuracy of CNN-based model is less than the VGG16 model, the proposed model surpasses the VGG16 in terms of complexity and execution time. Also, the proposed model shows a better performance compared to the Inception-resnet50 v2 and exception models.

Keywords: Deep learning, Transfer learning, AI, Healthcare.

1. Introduction

Tumor is an abnormal growth of human body cells. This growth comes in two kinds: benign tumor (non-cancerous), or malignant tumor (cancerous or incurable). The very sensitive type of tumor is the malignant brain tumor that could be developed inside brain tissue or speared from other body orgasms. The tumors that come from the brain cells or from the cells that cover the brain are called primary brain tumors. While the tumors that are transferred from other human-body orgasms are called metastatic tumors [1]. The malignant brain tumor leads to a serious damage to brain functionality or even death. To reduce the effect of this disease, many tumor identification techniques have been developed; the main aim of these techniques is to discover the tumor in its early stages to recover the brain from a curable damage. Magnetic resonance imaging (MRI) is an important technique because it is available in most regions, relatively affordable, and carries many features about the diagnosed brain areas [2]. Even though the usage of MRI is useful in diagnosing malignant tumors, doing the evaluation processes manually by a specialist is a challenging and time-consuming action. Moreover, there are at least 130 varying forms of brain-cell tumors that could be evaluated as benign or healthy tumors [3]. Therefore, image processing and image enhancement using Artificial intelligence (AI) techniques are adopted to facilitate the detection process of malignant-brain tumors. Artificial Intelligence algorithms have been used in the healthcare field for treatment and diagnosis since 1970 when the Stanford University developed a diagnosing system for blood diseases [4]. The race to develop efficient AI algorithms such as deep learning techniques led to present a list of helpful healthcare applications [5]. Generally, AI algorithms are categorized into several approaches: supervised, unsupervised, and semivised algorithms.

This work is licensed under a <u>Creative Commons Attribution 4.0</u> International License

Supervised learning algorithms require supervision from a specialist through learning processes by labeling the training examples in advance [6,7,8]. The researchers in [9] adopted the Support Vector Machine SVM algorithm to build a diagnosing system for brain tumors (glioma grading) ranging from I to IV according to the growth and aggressiveness of the tumor. Moreover, many diagnosing systems are developed by using deep-learning-based algorithms such as ResNets [10], GoogleNet [11], and VGGNets [12].

The unsupervised learning algorithms categorize the unlabeled data samples based on their structure to discover the unseen pattern within data [13,14]. Mainly, these algorithms are used with data clustering, where the datasets are divided into groups based on a clustering criterion that gathers similar data samples in a separate group. The number of groups is set in advance [15]. Goswami et al. segmented the brain tissues into several groups of MR images using the kmean algorithm [16]. Also, their system diagnosed the image data samples using Self Organized Map (SOM). O. Ciceketal. in [17] adopted a U-Net structure that involves two paths, down-sampling and up-sampling, to extract the contextual features and to localize tumors accurately using the 3D extension. Semi-supervised algorithms are also used in the fields of healthcare, e.g., brain tumors. Since the labeled data are limited in general, this approach is adopted to estimate the classes of unlabeled data samples based on the labeled ones. Ge et al. [18] proposed a graph-based semi-learning system to label the data samples with no class name. Augmentation with the GAN algorithm was used in their framework to produce more data samples. Diffusion models are used in [19] to segment the medical images of brain tumors and to learn the corresponding representations visually. Generally, the efficiency of AI algorithms is determined (among others) by two factors: model structure and the number of observed examples that are used to train the adopted structure. Training processing with a huge amount of data samples leads to an increase in the processing time, which in turn becomes an energy-consuming issue. Moreover, a complex structure of high trainable parameters requires a complicated computing architecture. So, the main gap observed in the literature is the structure complexity with a high processing time. The research question of this work is "How to develop a model with high performance and less complexity". This question is answered by suggesting a new model structure that inherences the main features, competes for accuracy, and finally surpasses the structure simplicity of the other transfer learning models. This work aims to enhance system architecture, reduce energy consumption, and improve performance.

The presented work is implemented in three phases. First of all, 3000 of brain MRI images are augmented with the aim of getting a generalized dataset that is suitable for the next steps. Second, a transfer learning approach is used by adopting the weights of three pretrained models namely, VGG16, Xception, and Inception-resnet50. These models are modified by adding an input layer, two dense layers, and an output layer. Then, a comparison study is conducted among the pretrained models to investigate their behaviors in classifying brain-tumors samples as healthy or malignant. This comparison study ends with the fact that the modified VGG16 model outperforms the others. This work continues by suggesting a novel design of CNN-based model with relatively high accuracy and a low level of complexity compared to the VGG16. The rest of the paper is arranged as follows: section 2 introduces the work methodology. The proposed model is discussed in section three. Section 4 summarizes the whole work.

2. Work methodology

This work's dataset is a brain-tumor-based set adopted from the Kaggle site [20]. The data set has 3000 samples of brain CT images that consist of 1500 healthy samples and 1500 affected samples. Examples of normal and malignant MRI images are displayed in Figures 1, and 2 respectively. In the conducted work, the dataset is shuffled and divided into three subsets: a training set of 2100 samples, a testing set of 600 samples, and 300 samples of validation. A preprocessing stage is conducted to resize samples, rescale the pixel values, and use a zoom range (0.2) with a horizontal flip. Images from Kaggle come with arbitrary sizes. So, the first step of the preprocessing stage is to resize all samples to 224x224. Then, the samples are normalized to the range of [0,1] to achieve a common contribution of the total loss from each image. Finally, the samples are augmented with zoom range and horizontal flip to produce multi-levels of zoomed images that flipped with different angles.





2.1. Comparison study

This section compares three popular pre-trained models: Vgg16, inception-resent, and exception model. The University of Oxford developed Visual Geometry Group (VGG). While inception-resent and Exception were proposed by Google researchers. The comparison procedure depends on the study of learning curves, confusion matrices, and evaluation metrics. Four evaluation metrics are adopted i.e., accuracy, precision, recall, and f1 score.

2.2. Comparison of learning curves

The learning curves of the three transferred learning models and the proposed model are illustrated in Figure 3. Each model is trained based on 100 iterations (epochs) using 2100 training samples. Within each epoch, a model is evaluated with an evaluation set (the 300 samples) to correct the training procedure based on the loss factor. The VGG16 model (red dot line) shows the best performance in terms of learning accuracy compared to other models. This comes, in part, due to the less complexity of VGG16 compared to Xception and Inception-resnet50 models. The complexity is measured (from our point of view) in terms of the number of trainable parameters. Where the VGG16 has 803410 parameters compared to Xception, and Inception-resnet50 v2 (3211858, and (54,276,192)) respectively.



2.3. Comparison of confusion matrices

In this section, a comparison among pretrained models is presented in terms of confusion matrices. a confusion matrix presents the model's behavior by showing the predicted labels compared to the original labels for each class in the dataset. It uses the true labels from the data sets to differentiate the image classes whether it is a health (H) or malignant (M) class. In this work, the dataset is divided into three subsets training, testing, and validation subsets with 2100, 600, and 300 brain-based samples. Table 1 depicts the true positive (malignant samples predicted as malignant), the true negative (healthy samples predicted as malignant), and the false negative (malignant samples predicted as healthy) for three pretrained models. Each model is evaluated with the three datasets. For example, the models' behavior with the training dataset shows the superiority of the VGG16 in terms of false negatives. False negative reads 4, 314, and 54 for VGG16, Xception, and Inception-resnet50 respectively. The VGG16 model keeps its superior behavior in terms of false negatives with both testing and validation subsets.

		VGG16		Xception			Inception-Resnet			
		Predicted			Predicted			Predicted		
ß	I		н	М		Н	М		H	М
raini	Actua	Η	1050	0	н	716	334	Н	830	220
I		М	4	1046	М	314	736	М	54	996
		Predicted		Predicted			Predicted			
esting	Actual		н	М		Н	М		Н	М
		H	300	0	H	215	85	H	233	67
L		М	2	298	М	91	209	М	6	294
		Predicted		Predicted		Predicted				
Validation	Actual		н	М		Н	М		Н	М
		Η	148	2	Н	116	34	Η	124	26
	7	М	6	144	М	47	103	М	4	146

2.4. Comparison of evaluation metrics

To test the predictability of the deep learning models, a testing dataset of 600 samples is used for evaluation. The predicted labels are evaluated with four metrices Accuracy, Precision, Recall, and F1 score. Next is the theoretical analysis of these metrics. Accuracy is calculated by dividing the number of truly predicted labels, (M,M) + (H,H) from the table (1), to the total number of labels as shown in Equation (1). It presents the overall behavior of a model. However, it does not illustrate the model behavior regarding each class in the data examples.

$$Accuracy = \frac{(M, M) + (H, H)}{Total number of samples}$$
(1)

Equation (2) calculates the Precision metric to estimate the percentage of positive samples that are predicted truly, (M, M), against the total number of samples that are predicted as positive, ((M,M)+(H,M)). Moreover, the Recall metric measures the rate of number of (M, M) to the total number of positive samples, ((M,M)+(M,H)), see Eq.(3). Finally, F1 score, Equation (4), is used to present the balance between Precision and recall.

$$Precision = \frac{(M,M)}{(M,M) + (H,M)}$$
(2)

$$Recall = \frac{(M, M)}{(M, M) + (M, H)}$$
(3)

$$F1\,Score = \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

Table (2) summarizes the behaviors of the three pretrained models based on the equations above. In the Xception model, precision, and recall of the malignant class are 0.71, and 0.7 respectively. Based on table (1), the 71% precision is obtained from the ratio of 209 (M, M) samples to 209+85, (M, M)+(H, M). The low precision of 0.71 comes from the fact that this

model falsely predicts 85 healthy samples as malignant, (false positive). On the other hand, this model reads 91 of malignant samples as healthy, false negative which leads to 70% of Recall based on equation (3). Therefore, the f1 score is 70% using eq (4). Additionally, the calculations of the healthy class are performed by considering the truly positive as (H, H) rather than (M, M) in Eq (1) and Eq (2).

Following the same calculation with VGG and Inception-Resnet, each value in Table (2) manifests a specific behavior aspect of a model with a particular class. For example, the precision of the healthy class in the Inception-resnet v2 model is 97% which is high despite the low accuracy (88%). This reflects that the Inception-resnet v2 model highly performs with healthy-class samples compared with the malignant class. By using the same analogy, the VGG16 model outperforms other models for both classes in terms of Precision and Recall, with 98% Precision for the healthy class, and 100% for the malignant class while 100% Recall for both classes is obtained. Therefore, VGG is considered the best choice as a pretrained model.

		Precision	Recall	F1-Score	Accuracy	Support
Xception	Healthy	0.7	0.72	0.71	0.71	600
	Malignant	0.71	0.7	0.7	0.71	
	Weighted Average	0.71	0.71	0.71	0.71	
Inception resnet v2	Healthy	0.97	0.78	0.86	0.88	600
	Malignant	0.81	0.98	0.89	0.88	
	Weighted Average	0.89	0.88	0.88	0.88	
	Healthy	0.00	1.00	1.00		600
Vgg16	Incarchy	0.55	1.00	1.00	1.00	
	Malignant	1.00	0.99	1.00		
	Weighted Average	1.00	1.00	1.00	1.00	

Table 2: Weighted evaluation metrics of three pretrained models

In conclusion, VGG16 shows a high performance compared to the other two models in terms of learning, confusion, and evaluation metrics. Here, the pretrained models are different in terms of the convolution layers; While the VGG 16 has thirteen CNN layers, the Xception has 36 CNN layers and Inseption-Resnet v2 exceeds this number. In fact, the reduction of CNN layers participates in reducing processing time, hardware complexity, and energy conception. Also, VGG16 has lower training parameters compared to other models which contributes to the reduction of the complexity. Therefore, proposing a new model with less complexity in terms of CNN layers and the number of trainable parameters with a high performance (compared with pretrained models) is the main aim of this study as discussed in the next section.

3. Proposed model

The research question of this work was "How to develop a model with high performance and less complexity". The answer to this question is presented in this section. A novel design of a CNN-based model with total trainable parameters of 71928 is suggested. The proposed model consists of an input layer, three convolutional layers, a flatten layer, two dense layers, and an output layer as shown in Figure (4).

The input layer reads a sample of a brain MRI image of size 224x224. The first CNN layer performs convolution operations with 64 filters of 3x3 dimensions. These operations are performed on an input sample of 222x222 size rather than 224x224 due to the neglecting of padding. The 224 is not a multiple of 3, so two pixels are neglected. The 64-feature maps of size 222x222 resulting from the first CNN layer are processed by a max-pooling layer with a window of size 2x2 to produce 64-feature maps of 111x111. The max-pooling layer focuses on the most important features within the window. Then, a second convolution layer is used to handle the (111x111) 64-feature maps of the first CNN-max-pooling layer with 32 filters of size 3x3 to produce 32-feature maps of size 109x109. Another max pooling layer is added to reduce the dimension of the feature maps of size 26x26. The feature maps are flattened to 10816 nodes. These nodes are attached to two dense layers of size 64, and 32 nodes. To generalize model behavior, two dropout layers are added after the second and third convolution layers. The Relu activation function is used with convolution and dense layers to evaluate the strong and weak signals. However, the Softmax activation function is used with the output layer to classify the input brain-based image to either affected (M) or healthy (H) patients as shown in Figure (4). This design leads to a reduction of the training parameters of 719281 compared to the pretrained-based models which contribute to reducing the complexity. Moreover, the performance is optimized by avoiding overfitting through the use of two dropout layers.



3.1. Model evaluation

The proposed model is trained with a brain-based image dataset of 3000 samples. These samples are divided into three subsets training, validation, and testing of 2100, 300, and 600 samples respectively. This section examines the performance of the suggested model through the same procedure followed with the evaluation of the pretrained models.

The first step in the examination process of this model is performed through the study of the learning curve. Figure (5) compares the learning curves of the suggested model with the three pretrained models studied earlier. It can be seen that our model outperforms the Xception and inception models concerning the epoch number, while it is still lower than that of VGG16.



Step 2 is performed through the confusion matrices of model behavior with three datasets (training, validation, and testing). This step aims to verify the model behavior by monitoring false positives (H, M) and false negatives (M, H) within all of the datasets as listed in Table (3).

During the training process, the proposed model is examined with 1050 healthy samples and the result shows that nineteen samples are falsely predicted as positive (H, M). While 28 samples are predicted falsely as negative (M, H) when the system is evaluated with 1050 malignant samples. Another round of verification is conducted using 300 samples labeled as validation dataset. With this dataset, the model falsely predicts eight of the healthy samples as positive (H, M). On the other side, three malignant cases are falsely predicted, i.e. (M, H).

Finally, with the testing dataset of 600 samples, the proposed model truly predicts 291 as (M, M) out of 300, while nine samples are falsely predicted as (M, H). On the other hand, this model has successfully recognized 286 of healthy samples, i.e. (H, H), out of 300 samples. The rest of the samples, (14 samples), are falsely recognized as malignant cases, (H, M). In the final step, the results obtained from confusion matrices in step two are evaluated using the four metrics discussed in

In the final step, the results obtained from confusion matrices in step two are evaluated using the four metrics discussed in section 2.2.3 and the evaluation scores are listed in Table (4).

		Proposed model					
		Predicted					
ing	al		Н	М			
rain	Actu	Н	1031	19			
T		Μ	28	1022			
		Predicted					
50	Ч		Н	М			
estin	Actua	Н	286	14			
L		М	9	291			
		Predicted					
ion	I		Н	М			
lidat	Actua	Н	142	8			
Val		М	3	147			

Table 3: Confusion matrices of the proposed model

Table 4: The evaluation scores of the proposed model

	Precision	Recall	F1-Score	Accuracy	Support	
Healthy	0.97	0.95	0.96	0.96		
Malignant	0.95	0.97	0.96	0.96	600	
Weighted Average	0.96	0.96	0.96	0.96		

4. Discussion and summary

The brain tumor is one of the most complicated diseases that requires an accurate and fast detection mechanism. MRI images are the common and most affordable way that provides an initial view of brain tissues. A human-based MRI detection process is prone to the possibility of a specialist mistake. Which in turn leads to a false diagnosis. AI provides a good contrivance to facilitate specialist diagnostics tasks. Different AI models have been used to classify brain-based MRI images. In this work, the weights of three pretrained models (namely, VGG16, Xception, and Inception-resnet50) are adopted to develop three brain-tumor detection systems. These systems are evaluated with 3000 MRI samples. VGG16 shows better performance in terms of the evaluation metrics (precession, recall, f1-score, and accuracy). However, the main issue with these models is the structure complexity. The complexity is compared in terms of trainable parameters and the number of CNN layers. Therefore, a low-complex CNN-based model is proposed to effectively detect the brain tumor in a relatively short time.

5. Conclusion and future work

The structure of the proposed model consists of three CNN-MaxPooling layers and a total of 719281 trainable parameters. This structure effectively reduces the complexity compared to the pretrained models. Where 803410, 3211858, and (54,276,192) are the trainable parameters of VGG16, Xception, and Inception-resnet respectively. The performance of the proposed model is evaluated and compared with the pretrained models in terms of evaluation metrics. The weighted evaluation metrics of our model are 0.96 which is higher than that of Xception, and Inception-resnet. However, it is 4% lower than that of VGG16 model. While the accuracy of our model is less than VGG16 model, the proposed model surpasses the VGG16 in terms of complexity and execution time.

In the near future, more transfer-learning models will be considered in the study plan to improve model capacity. Moreover, collecting and merging more tumor-based datasets is necessary to build a big dataset. This big dataset could add more features that help in identifying malicious tumors

References

- [1] Sapra P., Singh R., Khurana S. Brain tumor detection using neural network. Int. J. Sci. Mod. Eng. 2013;1:2319–6386.
- [2] Saxena, P., Maheshwari, A., Maheshwari, S. (2021). Predictive Modeling of Brain Tumor: A Deep Learning Approach. In: Sharma, M.K., Dhaka, V.S., Perumal, T., Dey, N., Tavares, J.M.R.S. (eds) Innovations in Computational Intelligence and Computer Vision. Advances in Intelligent Systems and Computing, vol 1189. Springer, Singapore. <u>https://doi.org/10.1007/978-981-15-6067-5_30</u>
- [3] Gore D.V., Deshpande V. Comparative study of various techniques using deep Learning for brain tumor detection; Proceedings of the 2020 IEEE International Conference for Emerging Technology (INCET); Belgaum, India. 5–7 June 2020; pp. 1–4.
- [4] Bush J. How AI is taking the scut work out of health care. Harvard Business Review 2018. https://hbr.org/2018/03/how-ai-is-taking-the-scut-work-out-of-health-care.
- [5] Cè M, Irmici G, Foschini C, Danesini GM, Falsitta LV, Serio ML, Fontana A, Martinenghi C, Oliva G, Cellina M. Artificial Intelligence in Brain Tumor Imaging: A Step toward Personalized Medicine. Curr Oncol. 2023 Feb 22;30(3):2673-2701. doi: 10.3390/curroncol30030203. PMID: 36975416; PMCID: PMC10047107.
- [6] Shalev-Shwartz S., Ben-David S. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press; Cambridge, UK: 2014.
- [7] Sidey-Gibbons J.A.M., Sidey-Gibbons C.J. Machine Learning in Medicine: A Practical Introduction. BMC Med. Res. Methodol. 2019;19:64. doi: 10.1186/s12874-019-0681-4.
- [8] Guido S., Muller A. Introduction to Machine Learning with Python a Guide for Data Scientists. O'Reilly Media; Sebastopol, CA, USA: 2018.
- K.E.Emblemetal., Predictive modelingingliomagrading from MR perfusion images using support vector machines. Mag Reson. Med.: Off. J. Int. Soc. Mag. Resona. Med. 60(4), 945–952 (2008)
- [10] He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition; Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Las Vegas, NV, USA. 27–30 June 2016; pp. 770–778.
- [11] Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. Going deeper with convolutions; Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); Boston, MA, USA. 7–12 June 2015; pp. 1–9.
- [12] Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv. 20141409.1556.
- [13] Jung A. Machine Learning: The Basics.Springer; Singapore: 2022.
- [14] Guido S., Muller A. Introduction to Machine Learning with Python a Guide for Data Scientists. O'Reilly Media; Sebastopol, CA, USA: 2018.
- [15] Cè M, Irmici G, Foschini C, Danesini GM, Falsitta LV, Serio ML, Fontana A, Martinenghi C, Oliva G, Cellina M. Artificial Intelligence in Brain Tumor Imaging: A Step toward Personalized Medicine. Curr Oncol. 2023 Feb 22;30(3):2673-2701. doi: 10.3390/curroncol30030203. PMID: 36975416; PMCID: PMC10047107.
- [16] S. Goswami and L. K. P. Bhaiya, "Brain Tumour Detection Using Unsupervised Learning Based Neural Network," 2013 International Conference on Communication Systems and Network Technologies, Gwalior, India, 2013, pp. 573-577, doi: 10.1109/CSNT.2013.123.
- [17] O. Ciceketal., 3D U-Net: learning dense volumetric segmentation from sparse annotation, in International Conference on Medical Image Computing and Computer-Assisted Intervention (Springer, Cham, 2016)
- [18] Ge, C., Gu, I.YH., Jakola, A.S. et al. Deep semi-supervised learning for brain tumor classification. BMC Med Imaging 20, 87 (2020). https://doi.org/10.1186/s12880-020-00485-0
- [19] Alshenoudy, A., Sabrowsky-Hirsch, B., Thumfart, S., Giretzlehner, M., Kobler, E. (2023). Semi-supervised Brain Tumor Segmentation Using Diffusion Models. In: Maglogiannis, I., Iliadis, L., MacIntyre, J., Dominguez, M. (eds) Artificial Intelligence Applications and Innovations. AIAI 2023. IFIP Advances in Information and Communication Technology, vol 675. Springer, Cham. <u>https://doi.org/10.1007/978-3-031-34111-3_27</u>
- [20] https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection?resource=download